A Visualization Approach to Dealing with Log Data

Emile Morse & Michelle Potts Steves

National Institute of Standards & Technology
100 Bureau Drive
Gaithersburg, MD 20899
(301) 975-8239
{emile.morse, msteves}@nist.gov

While conducting a usability study involving a team of geographically-dispersed welding researchers, it became apparent that the quantity of logged data was too large to handle with a solely statistical approach. The participants were being supported with a commercially available groupware environment that had been instrumented to allow the capture of user interaction information. Since visualization methods are considered to have potential value when large amounts of complex data are being treated, we developed a visual tool which we call the CollabLogger [2]. The interface allows the usability engineer to see an overview of the data and to select attributes for closeup analysis. When used in combination with standard spreadsheet and statistical methods, the visual tool has had its greatest value in allowing the usability professional to generate hypotheses. We believe that adding visual tools to the analytical toolkit will support better analyses of user behavior. In addition, the fact that team work produces logs that are not just quantitatively larger but more complex due to interactions among the test subjects, the utility of visual methods is probably greater than in simpler, single-subject situations.

1. Group Description

The core welding research team is comprised of five people with six roles divided among them and additional guest researchers, who may be distantly located from the physical location of the welding testbed itself. Each of the participants in the welding collaboratory has access to a Teamwave Workplace[1] (TW) client from his desktop computer as well as computers in the welding testbed. TW is a shared, room-based collaborative system with a WYSIWIS (what you see is what I see) whiteboard backdrop. Rooms provide boundaries for data groupings and user interactions and a metaphor for easing the transition in groupware. Doorways provide portals to other rooms. Data organization within rooms is configurable by its occupants in how they organize various tools housing their data, such as file viewers, PostIt™ notes, and message boards. FileHolders, and ImageHolders are used for linking documents and graphics that may have been created outside of TW . The system provides for synchronous and asynchronous user interactions, but importantly these interactions are in the context of relevant data.

The work in this experiment was characterized by several full group meetings (for planning and coordination) interspersed with periods of individual activity (asychronous work) and smaller coordination meetings of two or three team members around the “hand-off” of one task’s output to another task’s input.

2. Collected Data

The native version of TW produces logs on the server that contain information about the identity of users entering the distributed application, the identity of the rooms through which a user navigated, file uploads and message passing. This set of interactions was deemed to be too rudimentary for capturing the type of data that would be useful for usability analysis. Since the source code for the application was not available, we contracted with the vendor to instrument the software to place the appropriate hooks. Issues surfaced during the subsequent negotiation that are pertinent to the workshop focus. For example, the usability analysts tended to want more interactions logged than the vendor was willing to incorporate. In addition, logging at a fine-grain granularity had the potential to impact system performance. The outcome of the process was the creation of a second log which contained additional information such as the content of chat sessions, the combination of user ID, room ID and tool ID when a new tool was created, and the combination of user ID, tool ID and appropriate annotations when a tool was changed (e.g., message added to a Message Board or content of a PostIt modified).

The information automatically collected in logs was supplemented by questionnaires and interviews. The latter data is clearly qualitative, while log data has a more quantitative nature. Over the course of the study, which

[1] Teamwave Workplace [1] is a commercial product identified in this document for the purpose of describing a collaborative software environment. This identification does not imply any recommendation or endorsement by the National Institute of Standards and Technology.

lasted about 2 months, 4015 interactions were appended to the custom log file and 3605 events were created in the base server log.

3. Background on Time-based Visualizations

Data-rich environments are obvious targets for visual solutions. The following few paragraphs describe some visual approaches to handling time-based data.

LifeLines was developed by Plaisant et al [3] as a mechanism for displaying personal histories. They describe medical and juvenile justice scenarios implemented with the interface. Overview, zooming, and filtering are all facilitated in the LifeLines display. The data comprises information about single individuals and it is not apparent how the interface could be extended to cope with data from collaborative settings with many individuals.

Morse & Spring [4] presented a set of visualizations that were based on data gathered using the CASCADE (Computer Augmented Support for Collaborative Authoring and Document Editing) system. Although the environment was collaborative in nature, the data in the visualizations that they describe were filtered by users before rendering. This filtering leads to a situation in which relationships among team members can not be detected.

The Multi-Modal Logger (MML) was developed at MITRE [5] to record, retrieve, annotate and visualize data from a variety of sources, including audio and video. The visualization tool that is part of the MML toolkit presents the usability engineer with a timeline on the x-axis and a set of objects along the y-axis. Objects may represent users, audio logs, collaborative tool invocations or other significant components of the team environment. Filtering the data to select only the users yields a display that indicates when each user was active (i.e., creating loggable events) but makes it impossible to determine what the users were doing. On the other hand, if the display is not filtered then inferences about users' activities are possible but are made difficult due to the potentially large number of objects (i.e., timelines).

While log data for collaborative activities must be gathered at sufficient granularity to support all possible uses that might be made of the data, it is important to provide mechanisms at the interface to accomplish dynamic filtering of the data. The work on visualizations validates this approach to the treatment of time-based data, but also points out that none of the interfaces developed thus far has provided a full answer to the problems of handling multi-user, interaction data.

4. CollabLogger Visualization

The CollabLogger (Logger) application (Figure 1), implemented in Java, is different from those discussed above because it handles multiuser data, allows the usability engineer to focus on attributes of the data using a series of selection devices, and has a user-centric focus. The menu bar provides access to specialized functions such as replay and symbol overlay. The leftmost panel (A) provides labels for the timelines shown in the main panel of the display. These labels can be toggled between a user-centric view (shown in the Figure) and a tool-centric one. The figure shows log data zoomed in to a 4-hour time slice. The toggle button for user/tool mode is located immediately above the label list. The other button is a toggle that allows an overview of the entire log or a view of a single day.

The lower portion of the Logger window contains the controls by which the analyst interacts with the display. Checkboxes provide a mechanism to show or hide data in a particular view. Each room is color-coded and the panel C displays the mapping for all rooms in the log regardless of whether the current view has information about a particular room. The panel D shows the list of rooms that were occupied during the current day. Panel E allows selection of a subset of users. Information in the panels F and G is related to tool ID. Each instance of a tool is associated with a unique ID and these are displayed in panel F. Panel G allows hiding of tool information at the level of type of tool. For instance, the analyst may want to view all invocations of the 'Address Book' tool only. Events are shown in panel H and the user can delimit the set of markers by checking boxes next to the event type. Panel I allows selection among time segments based on days found in the current master log. Panel J provides a crude zoom tool.

The main display panel (B) of the figure shows that 5 sessions occurred during the time period under investigation. The bars show which room(s) each user occupied during his/her session. Tick marks colored to map to an event type are overlaid on these bars. The particular time period shown here reveals that more than one worker was present in the collaborative environment during most of the time. It also shows that these users

engaged in multiple chat sessions. By filtering the data with the controls, it is possible to explore the information space. Inferences can be drawn and hypotheses can be formed that might be subjected to statistical treatment.

Usability analysts working on collaborative data sets might be interested in answering the following kinds of questions:

- How often do multiple users occupy the space simultaneously?
- Which rooms are the most frequently visited?
- Are there patterns in the paths that users take within a set of rooms?
- Do particular users appear to be leading the activities?

Although the answers to the above questions could certainly be answered definitively using a statistical approach, visualizations allow the formulation of the questions themselves. Detecting patterns, outliers, densities, and gaps are innate abilities of the human visual system. A similar set of analyses can be performed in tool mode.

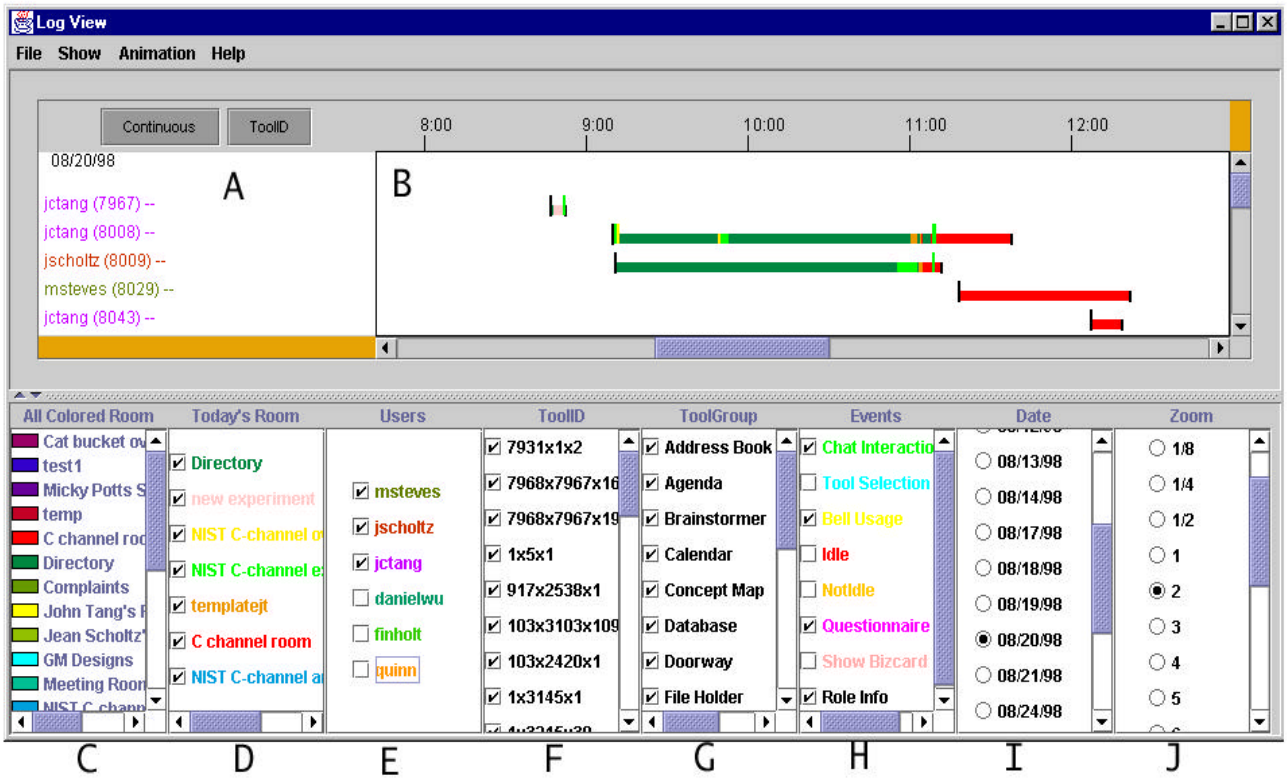

Figure 1: CollabLogger in user mode

5. Summary and Issues

Usability analyses of collaborative environments are difficult to perform partly due to the potentially complex interactions among the users of the groupware. Users interact not only with each other but also with information objects and tools provided by the environment. The Logger has been useful in allowing usability professionals to detect subtle interactions, which can then be subjected to more quantitative methods of analysis. One example of a use for the Logger is in characterizing asynchronous interactions. That is, how can you tell when people have created something in a room with the intent to communicate to others later on? The onset of a chat or of desktop conferencing is easily flagged as a synchronous episode, but problems arise when trying to make inferences about the contributions of asynchronous activity to the overall workflow. The logging tool supports detection of trails of tool usage that might prove to be sufficient to infer asynchronous communication.

Use of the Logger has provided evidence that visualization can be applied to the complex problem of mining log files. However, our use of the tool has also pointed out that it is limited when used alone, but derives power when used in combination with statistical analysis and examination of concurrent qualitative data.

[1] Roseman, M. 2000. http://www.teamwave.com/

[2] Morse, E., & M. Steves. 2000. CollabLogger: A Tool for Visualizing Groups at Work. WETICE 2000, 104-109

[3] Plaisant, C., B. Milash, A. Rose, S. Widoff, and B. Shneiderman. 1996. LifeLines: Visualizing Personal Histories. CHI 96, 221-227.

[4] Morse, E.L. and M.B. Spring. 1996. Visualizations that Support Group Work. IEEE Symposium on Visual Languages, Boulder, CO, September 3-6, 1996.

[5] MITRE Multi-Modal Logger. 1997. http://www.mitre.org/technology/logger/